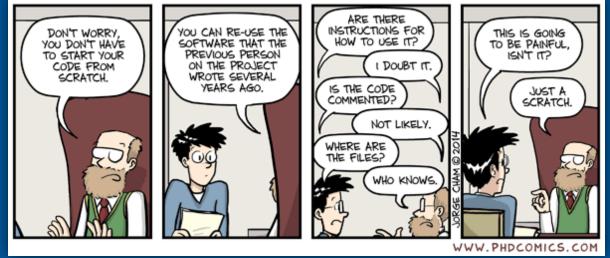
# Integrierte Entwicklungs- und Publikationsumgebung für Forschungssoftware und Daten am Helmholtz-Zentrum Dresden-Rossendorf (HZDR)

<u>Tobias Frust</u> Guido Juckeland Uwe Konrad





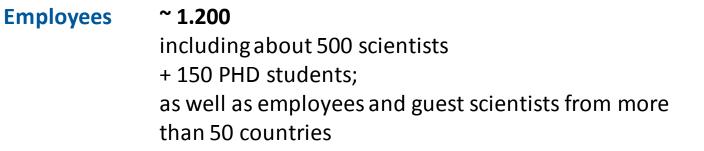


HELMHOLTZ | ZENTRUM DRESDEN | ROSSENDORF

HZD

#### **HZDR – Facts and Figures**

- Member of the Helmholtz Association
- Foundation 01.01.1992 (e.V.)
  Forschungszentrum Rossendorf







#### Research Sites Dresden

Helmholtz-Institut **Freiberg** Forschungsstelle **Leipzig** HIBEF-Station am XFEL **Schenefeld** Rossendorf Beamline an der ESRF in **Grenoble** 



Bilder: Killig, DESY, ESRF/Ginter





Member of the Helmholtz Association Tobias Frust | Department of Information Services and Computing | Computational Science Group | www.hzdr.de

#### Large Research Infrastructures

#### **ELBE - Center for High-Power Radiation Sources**

- Electron accelerator ELBE feeds free-electron lasers FELBE & THz source TELBE;
- generates positrons, protons and neutrons as well as X-ray and gamma radiation;
- plus high-intensity lasers (1 Petawatt) DRACO and PENELOPE (under construction)

#### **Dresden High Magnetic Field Laboratory (HLD)**

Producing Europe's highest pulsed magnetic fields for materials research

#### Ion Beam Center (IBC)

Nanoscale surface analysis and modification







Bilder: Bierstedt, Killig (2 x)





### **Research data and software – Where do you store it?**



?



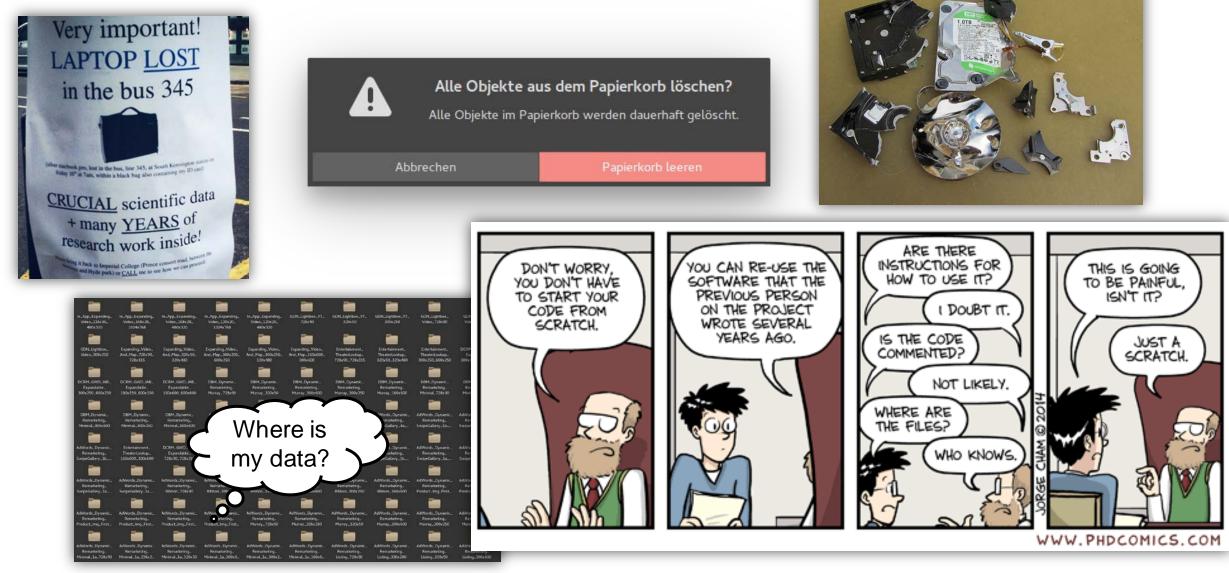






Member of the Helmholtz Association

#### **Research data and software – Is it safe?**



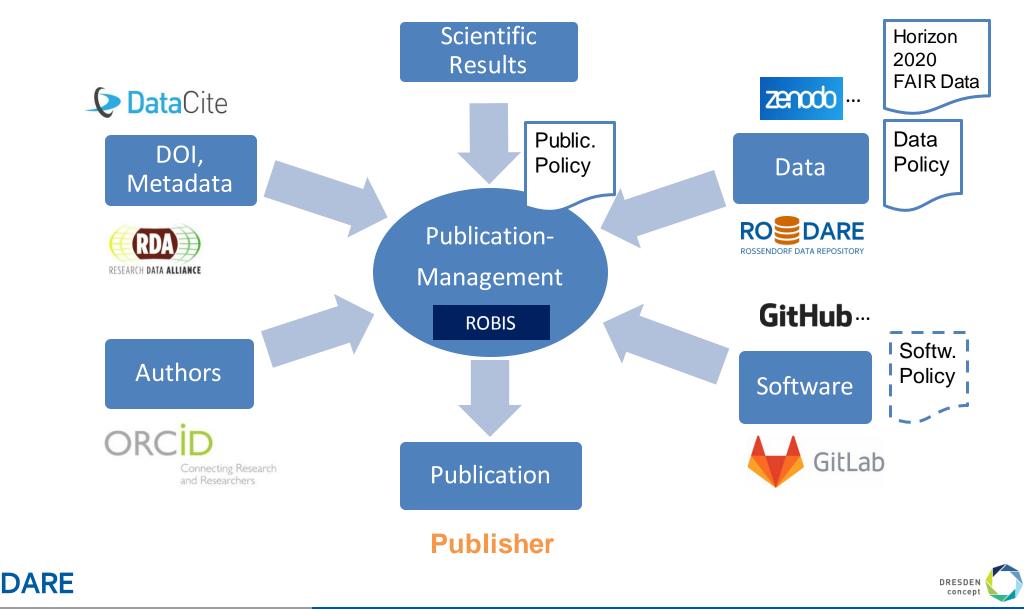


Member of the Helmholtz Association

DRESDEN

HZDR

### **Publication Components**



Member of the Helmholtz Association

HZDR

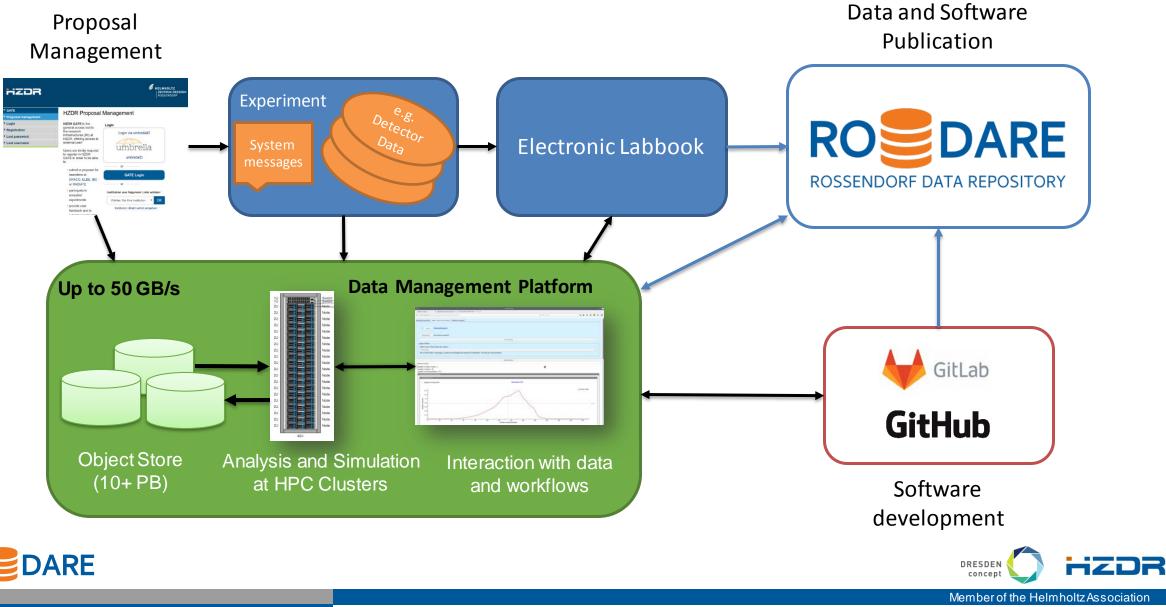
Tobias Frust | Department of Information Services and Computing | Computational Science Group | www.hzdr.de

RO

### Services for the whole lifecycle of data and software

RO

Page 7



#### **Research data and software – How to do it right?**



https://rodare.hzdr.de

| Recent uploads      March 9, 2019 (v0.8.0-slgha)    Software    Open Access      C++ & Python API for Scientific I/O with openPMD    View      © Koller, Fabian:    © Poeschel, Franz;    Huebl, Axel   | RODARE Docs<br>Have a look at the restructured<br>documentation and blog system of   |
|---|--|
| openPMD is an open metadata format for open data workflows in open science. This library provides a common high-level<br>API for openPMD writing and reading. It provides a common interface to I/O libraries and file formats such as HDF5 and<br>ADIOS. Where supported, openPMD-api implements both serial | RODARE. We now can more easily notify about<br>news and features. You also find tutorials there.<br>Visit https://rodare.hzdr.de/about.  |
| Uploaded on March 9, 2019<br>11 more version(s) exist for this record   | RODARE now offers usage statistics!  |
| March 6, 2019 (v1) Dataset Open Access View View Section of Pu(242) measured at the neutron time-<br>of-flight facility nELBE Kögler, Toni<br>This dataset includes the processed data of the fast neutron-induced fission of Pu(242) experiement performed in  | Thanks to the great folks<br>@inveniosoftware we are able to provide usage<br>statistics for record views and downloads.<br>Read the blog post to get more information about<br>the new feature. |
| November 2014 at the neutron time-of-flight facility nELBE which was published in T. Kögler et al., Phys. Rev. C 99, 024604<br>It contains the absolute (Pu242_nfis_Koegler_2019.csv) and   |  |
| Uploaded on March 7, 2019   | ROSDARE  |
| March 6, 2019 (vsubmission) Dataset Open Access View  | ROSSENDORF DATA REPOSITORY   |
| Supplementary Data: Spectral Control via Multi-Species Effects in PW-Class Laser-Ion<br>Acceleration  | Welcome to Rodare!   |
| 📀 Huebl, Axei; 😋 Rehwald, Martin; 📀 Obst-Huebl, Lieselotte; 📀 Ziegler, Tim; 💿 Garten, Marco; 🕤 Widera, René; Zeil, Karl;<br>💿 Cowan, Thomas E ; 💿 Bussmann, Michael; 💿 Schramm, Ulrich; 🌚 Kluge, Thomas   | The new data publication platform at HZDR.<br>Read more about Rodare on our  |

#### 

#### Why not simply Zenodo?

- Filesize limit 50 GiB; Dataset size limit 100 GiB
- → Not enough for the datasets we need to publish
- → Even if we had infinite space, bandwidth limitation
- Custom integrations for local experiments are planned



### **Challenges we are facing at HZDR**

- Many users don't see the need for open science yet.
- → Conviction process is required takes time. Journal requirements regarding open data/software often help.
- Publishing research data or software is often seen as only an additional burden.
- → Reduce the barrier by providing good integrations and establish obligatory guidelines.
- Knowledge about Open Science and "How to do it right" is often not available.
- $\rightarrow$  Training and support is required.



Photo by Mārtiņš Zemlickis on Unsplash





Page 9

#### Integrations – Background Upload

- Users need to be able to upload huge files (> 100 GiB)
  web upload is error prone.
- Many files already reside on central storage servers but under the control of the user.
- → Background upload via SFTP
- $\rightarrow$  Completely controllable via Web UI or Rest API

| SFTP  |                              |       | URL                       |
|---|------------------------------|-------|---------------------------|
| odare allows you to upload files via SFTI<br>a one of the registered servers you can t<br>earn more about the background upload | trigger an upload via the fi |       |                           |
| fes.hzdr.de   |                              |       | *                         |
| Q Filter by name  |                              |       | 15 elements per page 🔹    |
|   |                              |       |                           |
| 🖿 .ssh  |                              |       |                           |
| Documents   |                              |       |                           |
| 🖿 git   |                              |       |                           |
| 🖿 mozilla   |                              |       |                           |
| presentations   |                              |       |                           |
| Sicherung   |                              |       |                           |
| 🖿 test  |                              |       |                           |
| windows   |                              |       |                           |
| 🗅 .bash_history   |                              |       | (414 Bytes) 🗅 Upload file |
| 🗋 .gitconfig  |                              |       | 59 Bytes 🗅 Upload file    |
| 🗋 2014-15.pdf   |                              |       | 716.3 KiB 🗅 Upload file   |
| 🗋 2015-16.pdf   |                              |       | 799.1 KiB 🛆 Upload file   |
| 🗋 Bild1.emf   |                              |       | 4.9 MiB 🛆 Upload file     |
| 🗅 Bild1.wmf   |                              |       | 6.2 MiB  Upload file      |
|   | < 1 2                        | 3 4 > |                           |



Page 10



Member of the Helmholtz Association

#### **Integrations - GitLab**

- Publishing software releases should be as easy and as automated as possible.
- GitHub integration for Invenio already exists.
- A lot of institutions self-host a GitLab instance or use Gitlab.com, and so is HZDR.
- → Integration for private GitLab instances or GitLab.com is a popular request.
- → GitLab-Integration for Invenio: <u>https://gitlab.hzdr.de/rodare/invenio-gitlab</u>
- → Primary developed for Rodare; will hopefully be integrated into Zenodo,





Page 11





Member of the Helmholtz Association Tobias Frust | Department of Information Services and Computing | Computational Science Group | www.hzdr.de

### **Policies**

- HZDR Data Policy (since May, 2018). Instructions for:
  - Responsibilities
  - Data Management Plan (DMP)
  - Embargo period
  - Licensing of research data
- Software Policy
  - Under construction in collaboration with the Helmholtz Task
    Group Open Science
  - First draft of a template is now discussed with other responsibilities (e.g. Technology Transfer, Legal department, ...)

→ Data Librarian: Provide support and training for scientists







## Invenio RDM [1] – The Future of RODARE

The project has an ambitious one year schedule to build an RDM solution based on Zenodo. Invenio RDM will deliver:

- Invenio RDM A research data management platform based on Zenodo and Invenio v3.
- A community of public and private institutions to sustain Invenio RDM.
- Minimum two existing repositories migrated to Invenio RDM, with Zenodo being one of them.











Member of the Helmholtz Association

DARE [1] <u>https://invenio-software.org/blog/2019-04-29-rdm/</u>

- → Core repository: extensible metadata model based DataCite metadata schema with support for handling millions of records and peta bytes of data
- → Packaging and distribution: require minimal experience in installing, operating and administering the platform

[1] https://invenio-software.org/blog/2019-04-29-rdm/

→ Customization and extendability: easy extensibility and customisability to adapt to each particular institution





INVENIO) RDM





Member of the Helmholtz Association

Tobias Frust | Department of Information Services and Computing | Computational Science Group | www.hzdr.de

Page 14

DARE

#### Conclusion

- Publication of research data and software requires integrated services.
- Metadata and documentation should be captured continouosly and as soon as possible.
- Metadata should be collected as automated as possible.
- Users need adapted and individual support and training.

For the service provider:

- Installation, administration and deployment should be as easy as possible.
- Easy extensibility and customizability.
- Use widely accepted standards, e.g. for metadata formats.



Page 15



Thank you!

What do you think?

Are we doing data and software publication right?



Powered by



Member of the Helmholtz Association

Tobias Frust | Department of Information Services and Computing | Computational Science Group | www.hzdr.de



Page 16